

Final Exam 3**May 12, 2016****90 minutes**

Name: _____

Cell phones and any communication tools are not permitted. No interaction of any sort is allowed with your classmates.

Please show all your work, including all calculations, and explain your answers.

Part	Points Scored	Out of
1		20
2.1		10
2.2		10
Total		40

1 Election (20 points total)

Can we use state-level variables to predict whether a state votes for the Democratic versus Republican nominee? The `Stat2Data > Election08` data file contain 50 states plus the D.C. The variable recorded are:

- State: abbreviation for the state
- Income: per capita income as of 2007
- HS: percentage of adults with at least a high school education
- BA: percentage of adults with at least a college education
- Dem.Rep: percentage difference of the Democratic to Republican leanings in a state
- ObamaWin: An indicator of whether or not the Democratic candidate Barak Obama won a majority of votes in the state

1. (4 points) Fit separate logistic regression models to predict *ObamaWin* using each of the predictors *Income*, *HS*, *BA* and *Dem.Rep*. Which of these variables does the most effective job of predicting this response? Which is the least effective? Explain the criteria you use to make these decisions.

2. (3 points) Consider a logistic regression model to predict *ObamaWin* for each state using the per capita *Income*. The units of the *Income* is in \$1. In order to interpret the *Income* coefficient in a meaningful way, let's to re-express the income variable in \$1000. (a) How does the fitted prediction equation change? (b) Interpret the coefficient directly.

3. (4 points) Use the per-capita income of Massachusetts and New Hampshire directly to describe the odds ratio of *ObamaWin*. Show the calculation below, and interpret it.
4. (4 points) Answer one of the three: (a) Find a 90% confidence interval of an estimated odds ratio in 3. (b) Find a 90% confidence interval of an estimated *ObamaWin* percentage in 2008. (FYI, the actual percentage of votes was 61.80.)

2 Ethical Considerations in Data Analysis

Gather in a group of 4-6 people. There are two independent part to this discussion. First read the questions and mark **x** next to the question that you think you are most comfortable answering (yourself). After you answer the questions, please put a circle by the question that you wish I grade. There is space to take notes while you develop common ideas and language to respond to the questions.

2.1 How a Gay-Marriage Study Went Wrong

8. List two things **you** would have done to prevent disgrace in reliability in research publication, assuming you were the political scientist very much interested in how people view the issue of gay-marriage.

Here are the prompts you could use to generate discussion as a group. Jotting down some notes below for these questions may be helpful for you and me to understand your final suggestions.

- Summarize the process of the research.
- Why was LaCour tempted to fabricate results, you think? (What are the bad incentives that block conscientious research?)
- How would you want to review s research study–yourself, ask independent researcher/scholar, or trust the journal editors?

2.2 Science Isn't Borken

Some quotes from the article:

Which political party is best for the economy seems like a pretty straightforward question. But as you saw, its much easier to get a result than it is to get an answer.

The p-value reveals almost nothing about the strength of the evidence, yet a p-value of 0.05 has become the ticket to get into many journals...Instead of p-values, the journal will require **strong descriptive statistics, including effect sizes**... p-value as an index of surprise. How surprising would these results be if you assumed your hypothesis was false?

Here are the recommended steps for the group:

1. Let each person be responsible for summarizing one section (two-four paragraphs).
 2. Link the keywords and piece together what p-hacking means. You may look up the term online for 1 minute (but no more)!
 3. In the context of analyzing political economy (the first interactive tool on the webpage), would you have
9. Focus on the section directly below the picture of soccer-game and two-three graphs below. Describe in your own words, what p-hacking is. List two reasons why it is frowned upon. Be as statistically precise as you can (for example, in terminology choice or reasoning) when you give your reasons. For each reason, describe in detail where the serious concerns arise
- (a) Initial stages: goal of survey, design and data collection: Is the study being conducted to prove or reinforce a particular point of view, or is it to find the truth? Who is the party responsible for financing the study?
 - (b) Intermediary stages: data organization, analysis and presentation – the methods employed in the study
 - (c) Final stages: interpreting the outcome

10. An R-squared of 0.09 (i.e. correlation of 0.3) is not a miserable output when we examine the relationship between an economic variable, say employment rate, vs. the proportion of Democratic party/power (in the U.S). Discuss as a group, what the Pearson correlation coefficient means and how the numeric value of 0.3 explains how the human society is organized. How would you justify the statement above to your peers in a political science class?